Inference for Relationships Checkpoint

The following questions present you with a scenario, and you need to choose the most appropriate statistical test in each case.

Pool1

Question (1)

We suspect the overall mean monthly rent of apartments in Shadyside is higher than in Oakland, so we survey a random sample of *Oakland* apartments, and a random sample of *Shadyside* apartments.

A: matched pairs t-test
B: two-sample t-test
C: ANOVA
D: chi-squared test for independence
E: inference for regression

Feedback

A : 0

✗ This is not quite right. Notice that you want to test if the means of two random samples are equal. Consider the remaining options. (B) is the right answer.

B : 10

✓ Good job! We are comparing two groups—apartments in Shadyside and apartments in Oakland. Since we take *two random samples*, these samples are *independent* and therefore the two-sample t-test is appropriate.

C : 0

✗ This is not quite right. Notice that you want to test if the means of two random samples are equal. Consider the remaining options. (B) is the right answer.

D : 0

✗ This is not quite right. Notice that you want to test if the means of two random samples are equal. Consider the remaining options. (B) is the right answer.

E : 0

✗ This is not quite right. Notice that you want to test if the means of two random samples are equal. Consider the remaining options. (B) is the right answer.

Question (2)

At the beginning of the semester, students who registered for a statistics course were randomly assigned to two sections, each taught by a different instructor. At the end of the semester, we would like to test whether there are differences in performance on the final exam between the two sections.

A: matched pairs t-test
B: two-sample t-test
C: ANOVA
D: chi-squared test for independence
E: inference for regression

Feedback

A : 0

✗ That's not quite right. The students are randomly assigned to a section, and therefore, we have two independent samples. (B) is the right answer.

B : 10

✓ Good job! By randomly assigning students to a section, we've formed two independent samples. A two-sample t-test is therefore the appropriate test.

C : 0

✗ That's not quite right. Note that we're comparing only two independent samples. ANOVA is used when we're comparing more than two groups. (B) is the right answer.

D : 0

✗ That's not quite right. A chi-squared test for independence is used to determine if there is an association between two categorical variables. A grade on a final exam is quantitative. (B) is the right answer.

E : 0

✗ That's not quite right. A regression is used for exploring the relationship between two quantitative variables. Here we are interested in the differences in final exam scores between two independent samples. (B) is the right answer.

Pool2

Question (3)

We suspect that automobile insurance premiums (in dollars) may be steadily decreasing with the driver's driving experience (in years), so we choose a random sample of drivers who have similar automobile insurance coverage and collect data about their ages and insurance premiums.

A: matched pairs t-test
B: two-sample t-test
C: ANOVA
D: chi-squared test for independence
E: inference for regression

Feedback

A : 0

✗ This is not quite right. Notice that you are measuring two quantitative variables from each subject in order to see if there is a trend between them. Consider the remaining options. (E) is the right answer.

B : 0

✗ This is not quite right. Notice that you are measuring two quantitative variables from each subject in order to see if there is a trend between them. Consider the remaining options. (E) is the right answer.

C : 0

✗ This is not quite right. Notice that you are measuring two quantitative variables from each subject in order to see if there is a trend between them. Consider the remaining options. (E) is the right answer.

D : 0

✗ This is not quite right. Notice that you are measuring two quantitative variables from each subject in order to see if there is a trend between them. Consider the remaining options. (E) is the right answer.

E : 10

✓ Good job! We are examining the relationship between *two quantitative* variables (case Q&T'Q)—premium (in dollars) and *driving experience* (in years), and therefore inference on regression is appropriate.

Question (4)

Advertising researchers claim that the power of curiosity can be harnessed to design an effective Internet advertising strategy that results in a better evaluation of the advertised product. They develop six advertising texts with varying amounts of "curiosity" triggers. College students are randomly assigned to one of the six versions of the advertisement text and their evaluation score of the advertised product is recorded.

A: matched pairs t-test
B: two-sample t-test
C: ANOVA
D: chi-squared test for independence
E: inference for regression

Feedback

A : 0

✗ That is not quite right. Since students are randomly assigned to one of the six versions, there are six independent samples. (C) is the right answer.

B : 0

✗ That is not quite right. Note that we're comparing more than two independent samples. (C) is the right answer.

C : 10

✓ Good job! ANOVA is used for comparing more than two groups, and therefore can be used to determine if the mean evaluation score is the same for all six advertising texts.

D : 0

✗ That is not quite right. The chi-squared test is used for examining the relationship between two categorical variables.

(C) is the right answer.

E : 0

✗ That is not quite right. We are examining whether evaluation score is related to version, which is a categorical variable. Regression is used for examining a relationship between two quantitative variables. (C) is the right answer.

Pool3

Question (5)

To test whether Internet use increases depression score, we measure the depression scores of a random sample of non-Internet-users, have them use the Internet for a specified time, then measure their depression scores again.

A: matched pairs t-test
B: two-sample t-test
C: ANOVA
D: chi-squared test for independence
E: inference for regression

Feedback

A : 10

✓ Good job! Since we are measuring the depression scores of the same group of subjects, before and then after using the Internet, the matched pairs t-test is appropriate.

B : 0

✗ This is not quite right. Notice that one variable is being measured on each subject twice, once before using the Internet and once after using the Internet, to see the effect of Internet use on depression. Consider the remaining options. (A) is the right answer.

C : 0

✗ This is not quite right. Notice that one variable is being measured on each subject twice, once before using the Internet and once after using the Internet, to see the effect of Internet use on depression. Consider the remaining options. (A) is the right answer.

D : 0

✗ This is not quite right. Notice that one variable is being measured on each subject twice, once before using the Internet and once after using the Internet, to see the effect of Internet use on depression. Consider the remaining options. (A) is the right answer.

E : 0

✗ This is not quite right. Notice that one variable is being measured on each subject twice, once before using the Internet and once after using the Internet, to see the effect of Internet use on depression. Consider the remaining options. (A) is the right answer.

Question (6)

A physical therapy researcher was interested in determining the impact of two different exercises. The investigator suspected that the two exercises produced a different level of activity in the muscle. Each of 16 subjects performed both exercise 1 and exercise 2, and the results are shown (for each subject, the order of the exercises was randomly assigned and sufficient rest time was provided between the two exercises).

Subject	Exercise 1	Exercise 2
1	4.53	5.00
2	6.72	7.12
3	3.79	4.21
4	5.82	5.78
5	5.00	6.29
6	4.00	5.13
7	2.19	3.95
8	7.34	7.20
9	9.12	8.95
10	1.79	4.12
11	8.10	9.68
12	7.52	8.25
13	6.25	7.15
14	6.60	7.98
15	7.00	7.10
16	6.23	6.00

A: matched pairs t-test
B: two-sample t-test
C: ANOVA
D: chi-squared test for independence
E: inference for regression

Feedback

A : 10

✓ Good job! Indeed, we are comparing two groups in which each observation in one sample is linked to an observation in the second sample (the same subject is measured twice).

B : 0

✗ That is not quite right. While it is true that we are comparing two groups, the two samples are not independent, since each subject was measured twice. (A) is the right answer.

C : 0

✗ That is not quite right. Note that the samples here are not independent, and that we are comparing only two groups. (A) is the right answer.

D : 0

✗ That is not quite right. Chi-squared is used for two categorical variables. The score on the assessment is a quantitative variable. (A) is the right answer.

E : 0

✗ That is not quite right. Note that we are comparing two groups in which each observation in one sample is linked to an observation in the second sample. (A) is the right answer.

Pool4

Question (7)

We select random samples from several racial categories (Caucasian, African-American, Hispanic, Asian-American) to determine if there is a difference in overall mean earnings among the groups.

A: matched pairs t-test
B: two-sample t-test
C: ANOVA
D: chi-squared test for independence
E: inference for regression

Feedback

A : 0

✗ This is not quite right. Notice that you are testing if there is a difference in means between multiple groups. Consider the remaining options. (C) is the right answer.

B : 0

✗ This is not quite right. Notice that you are testing if there is a difference in means between multiple groups. Consider the remaining options. (C) is the right answer.

C : 10

✓ Good job! We are comparing the earnings of 4 racial groups and therefore the ANOVA F-test is appropriate.

D : 0

✗ This is not quite right. Notice that you are testing if there is a difference in means between multiple groups. Consider the remaining options. (C) is the right answer.

E : 0

✗ This is not quite right. Notice that you are testing if there is a difference in means between multiple groups. Consider the remaining options. (C) is the right answer.

Question (8)

Researchers question whether college students' choice of declared academic major is related to gender.

A: matched pairs t-test
B: two-sample t-test
C: ANOVA
D: chi-squared test for independence
E: inference for regression

Feedback

A : 0

✗ That is not quite right. Both gender and declared academic major are categorical variables. For a matched pairs t-test, the response variable must be quantitative. (D) is the right answer.

B : 0

✗ That is not quite right. Both gender and declared academic major are categorical variables. For a two-sample t-test, the response variable must be quantitative. (D) is the right answer.

C : 0

✗ That is not quite right. Both gender and declared academic major are categorical variables. For an ANOVA, the response variable must be quantitative. (D) is the right answer.

D : 10

✓ Good job! A chi-squared test for independence will determine if there is an association between the two categorical variables, gender and declared academic major.

E : 0

✗ That is not quite right. Gender and academic major are categorical variables. Regression is used for two quantitative variables. (D) is the right answer.

Pool5

Question (9)

We want to test for a relationship between race and marital status (married/never married/divorced/widowed).

A: matched pairs t-test
B: two-sample t-test
C: ANOVA
D: chi-squared test for independence
E: inference for regression

Feedback

A : 0

✗ This is not quite right. Notice that each subject falls into one race category as well as one marital status category and you want to test if there is a significant relationship between these two variables. Consider the remaining options. (D) is the right answer.

B : 0

✗ This is not quite right. Notice that each subject falls into one race category as well as one marital status category and you want to test if there is a significant relationship between these two variables. Consider the remaining options. (D) is the right answer.

C : 0

✗ This is not quite right. Notice that each subject falls into one race category as well as one marital status category and you want to test if there is a significant relationship between these two variables. Consider the remaining options. (D) is the right answer.

D : 10

✓ Good job! We are examining the relationship between two categorical variables—race and marital status (case Câ†'C) and therefore the appropriate inferential procedure is the chi-squared test for independence.

E : 0

✗ This is not quite right. Notice that each subject falls into one race category as well as one marital status category and you want to test if there is a significant relationship between these two variables. Consider the remaining options. (D) is the right answer.

Question (10)

We want to explore the relationship between the prices of diamond rings and the weights of their diamond stones.

A: matched pairs t-test
B: two-sample t-test
C: ANOVA
D: chi-squared test for independence
E: inference for regression

Feedback

A : 0

✗ That is not quite right. Note that we are not comparing two groups here, but rather examining the relationship between two quantitative variables. (E) is the right answer.

B : 0

✗ That is not quite right. Note that we are not comparing two groups here, but rather examining the relationship between two quantitative variables. (E) is the right answer.

C : 0

✗ That is not quite right. Note that we are not comparing several groups (which is what ANOVA is used for), but rather exploring the relationship between two quantitative variables. (E) is the right answer.

D : 0

✗ That is not quite right. Recall that the chi-squared test is used to explore the relationship between two categorical variables. Note that price and weight are two quantitative variables. (E) is the right answer.

E : 10

✓ Good job! Indeed, regression is used to examine the relationship between two quantitative variables such as price and weight.

Pool6

The next three questions refer to the following information:

Suppose an economist wishes to determine the relationship between the age and price of houses. A study yields the following data:

Worksheet 1 ***

	C1	C2
1	27	48
2	49	63
3	35	50
4	51	72
5	42	55

For each question below, choose the most appropriate inference method to analyze these data, under the given scenarios.

Question (11)

If column 1 is the price (in thousands of dollars) of a sample of five houses from ten years ago, and column 2 is the price (in thousands of dollars) of a sample of a different five houses from today, which of the following is the appropriate inference method?

A: matched pairs
B: two independent samples
C: inference for regression

Feedback

A : 0

✗ This is not quite right. Notice that you are given two independent samples from different times. Consider the remaining options. (B) is the right answer.

B : 10

✓ Good job! We are comparing the prices of houses in two time periods using two (different) independent samples.

C : 0

✗ This is not quite right. Notice that you are given two independent samples from different times. Consider the remaining options. (B) is the right answer.

Question (12)

If column 1 is the age of the home in years, and column 2 is the price of the home (in thousands of dollars), which of the following is the appropriate inference method?

A: matched pairs
B: two independent samples
C: inference for regression

Feedback

A : 0

✗ This is not quite right. Notice that each subject is measured

✗ once for age and a second time for price, and that you want to see if there is a trend between the age of a house and its price. Consider the remaining options. (C) is the right answer.

B : 0

✗ This is not quite right. Notice that each subject is measured once for age and a second time for price and that you want to see if there is a trend between the age of a house and its price. Consider the remaining options. (C) is the right answer.

C : 10

✓ Good job! We are examining the relationship between two quantitative variables—the age of the house and its price, and therefore inference on regression is appropriate.

Question (13)

If column 1 is the price (in thousands of dollars) of a sample of five houses from ten years ago, and column 2 is the price (in thousands of dollars) of the same homes today, which of the following is the appropriate inference method?

A: matched pairs
B: two independent samples
C: inference for regression

Feedback

A : 10

✓ Good job! We are comparing the prices of houses in two time periods using the same group of 5 houses measured twice, and therefore matched pairs is appropriate.

B : 0

✗ This is not quite right. Notice that each subject is measured twice, once in each time period, for its price and that you want to see if the mean prices of the houses have changed over time. Consider the remaining options. (A) is the right answer.

C : 0

✗ This is not quite right. Notice that each subject is measured twice, once in each time period, for its price and that you want to see if the mean prices of the houses have changed over time. Consider the remaining options. (A) is the right answer.

The next three questions refer to the following information:

A researcher wants to explore the relationship between the following two variables, Weight 1 and Weight 2.

Weight 1	Weight 2
130	138
160	156
220	224
125	119
205	213

For each question below, choose the most appropriate inference method to analyze these data, under the given scenarios.

Question (14)

Suppose that Weight 1 is the weight (in pounds) of a sample of five individuals before beginning a weight-loss diet, and Weight 2 is the weight in (pounds) of the ***same*** five individuals after the diet. If we would like to test the effectiveness of the diet, which of the following is the appropriate inference method?

A: matched pairs
B: two independent samples
C: inference for regression

Feedback

A : 10

✓ Good job! A matched pairs study may be carried out, where each observation in one sample is matched/paired/linked with an observation in the other sample. In this case, we have the pairs of before-diet and after-diet weights of each individual.

B : 0

✗ That is not quite right. The samples represent weights from the same five individuals. These are not independent samples. These are matched pairs. (A) is the right answer.

C : 0

✗ That is not quite right. Note that in this scenario, each observation in one sample is linked to an observation in the second sample, and we're exploring whether the weights after the diet tend to be lower than the weights before the diet. (A) is the right answer.

Question (15)

In preparing for a balsa wood bridge challenge, students weighed (in g) a random sample of five balsa wood designs and recorded these data as Weight 1. They also recorded the maximum weight (in kg) that the bridge could support. If the students want to know if there is an association between the weight of the bridge (in g) and the maximum weight supported (in kg), which of the following is the appropriate inference method?

A: matched pairs
B: two independent samples
C: inference for regression

Feedback

A : 0

✗ That is not quite right. Regression is the correct choice, since we're exploring the relationship between two quantitative variables. The students want to know if there is an association between the weight of the balsa wood bridge and the weight that the bridge supported. (C) is the right answer.

B : 0

✗ That is not quite right. Note that we're not comparing two groups here, but rather exploring the relationship between two quantitative variables, and therefore regression is the correct choice. Students want to know if there is an association between the weight of the balsa wood and the weight that the bridge supported. (C) is the right answer.

C : 10

✓ Good job! The students want to know if there is an association between the weight of the balsa wood and the weight that the bridge supported.

Question (16)

If Weight 1 is the weight (in pounds) of a random sample of five men who were accepted as models, and Weight 2 is the weight (in pounds) of a random sample of five men who were rejected as models, which of the following is the appropriate inference method?

A: matched pairs
B: two independent samples
C: inference for regression

Feedback

A : 0

✗ That is not quite right. The weights are from two independent samples. (B) is the right answer.

B : 10

✓ Good job! The weights are from two independent samples that we are comparing.

C : 0

✗ That is not quite right. The weights are from two independent samples. In regression, you are testing for an association between two variables, where each (x,y) pair are somehow related. There is no indication that the weight of accepted model 1 is in any way associated with the weight of rejected model 1. (B) is the right answer.